OCCUPATION AND CANCER IN ONTARIO:
REVIEW OF THE OPTIONS FOR ESTABLISHING A CANCER-OCCUPATION DATA BASE FOR ONTARIO

SUMMARY REPORT

Submitted by

Loraine D. Marrett and Erica Weir
Ontario Cancer Treatment and Research Foundation

Industrial Disease Standards Panel

Occasional Paper

May 24, 1989

91-01556/

Canadian Cataloguing in Publication Data

Marrett, L.D.

Occupation and cancer in Ontario--review of the options for establishing a cancer-occupation data base for Ontario

(IDSP Occasional paper)
ISBN 0-7729-7122-6

1. Occupational diseases--Ontario.

2. Information storage and retrieval systems --Carcinogens.

I. Weir, Erica.
II. Ontario. Industrial Disease Standards Panel.
III. Title.

RA645.C3M37 1990 616.9'803 C90-092529-9

OCCUPATION AND CANCER IN ONTARIO: REVIEW OF THE OPTIONS FOR ESTABLISHING A CANCER-OCCUPATION DATA BASE FOR ONTARIO

SUMMARY REPORT

TABLE OF CONTENTS

1. INTRODUCTION

2. METHODS

3. SUMMARY OF FINDINGS

- 3.1 The Ontario Cancer Registry (OCR)
- 3.2 Computerized Linkage of Administrative and Health Databases
- 3.3 Potentially Linkable Occupational Databases: a) The Census; b) Unemployment Insurance (UIC) Book Renewal Cards; c) Income Tax records; d) Canadian Labour Force Survey; e) Combined Employment Records.
- 3.4 Generation of New Occupation Data for Cancer Patients
- 3.5 Enhancement of Occupational Data
- 3.6 Exposure Databases

4. RECOMMENDATIONS

TABLE 1: DESCRIPTION OF POTENTIALLY LINKABLE OCCUPATIONAL DATABASES

TABLE 2: COMPARISON OF POPULATION-BASED STUDIES SYSTEMATICALLY COLLECTING OCCUPATION DATA ON NEWLY-DIAGNOSED CANCER PATIENTS

APPENDIX 1: TERMS OF REFERENCE FOR THIS REVIEW

APPENDIX 2: LIST OF CONTACTS MADE DURING THIS REVIEW

APPENDIX 3: LETTER FROM MR. BRUCE PETRIE, STATISTICS CANADA

APPENDIX 4: CANCER ACT OF ONTARIO (REVISED STATUTES, 1980)

APPENDIX 5: DESCRIPTION OF ONTARIO CANCER REGISTRY

OCCUPATION AND CANCER IN ONTARIO: REVIEW OF THE OPTIONS FOR ESTABLISHING A CANCER-OCCUPATION DATA BASE FOR ONTARIO SUMMARY REPORT

1. INTRODUCTION

Ontario has no systematic means for generating hypotheses linking cancer incidence and employment, industry, occupation or workplace exposures. This fact stands in the way of reliably estimating the "burden of occupational cancers" in this province.

This review was commissioned by the Industrial Disease Standards Panel. Its primary objective was "to determine whether an occupational cancer database can be developed (for the province) by linking occupational data with appropriately matched records of the Ontario Cancer Treatment and Research Foundation (OCTRF)". Appendix 1 contains the complete Terms of Reference. The existence of this database would enable quantification and description of the occurrence of occupational cancer in Ontario.

The source of cancer information would, of course, be the Ontario Cancer Registry (OCR) maintained by the OCTRF. The authors examined several sources for existing occupational data and different means for generating new data. They also explored approaches for linking cancer and occupational data. The strengths and limitations of each option are discussed in depth in the review. Factors such as cost, data quality, feasibility, accessibility, scientific validity, and statistical power were considered, and some rough estimates of time frames and staffing requirements provided.

The review focussed on two major approaches to the creation of an occupation-cancer data base:

1. Linkage of existing occupational data collected prior to a diagnosis of cancer as reported to the Ontario Cancer Registry: and
2. Collection of new occupational data on cancer patients reported to the OCR.

With respect to the first approach, only those sources of occupational data considered to be of high quality and for which linkage with the OCR was likely to be feasible are covered in depth. For the second approach, the authors have concentrated on sources of occupational information which could be readily gathered under current conditions (viz. the collection of occupational histories from cancer patients and/or population controls) rather than the prospective gathering of exposure data using the new WHMIS (for Workplace Hazardous Materials Information System) legislation.

The search for sources of data focussed on occupational, rather than exposure data, pursuant to the Terms of Reference. It would be desirable, indeed preferable, to have exposure information for all occupations, but at present such information exists only for a few select jobs. Expansion to all jobs would be a large and formidable task, which may be accomplished gradually under the new WHMIS legislation. The use of job-exposure matrices to impute exposures from job titles was also investigated.

The review, comprising several hundred pages of text, recommendations and supporting appendices, was submitted to the Panel in September, 1988. This paper summarizes the methods, findings and recommendations in that review.

2. METHODS

The authors reviewed relevant documents and literature, and held discussions with key individuals. The reviewed documents and literature are listed at the end of each section in the full report while the consulted individuals and their affiliations are shown in Appendix 2. Several names are highlighted below because of the importance of their contribution to this review.

At Statistics Canada, discussions were held with several staff on a variety of issues:

- John Silins and Martha Fair of the Health Division on computerized linkage of health and administrative databases;
- Bruce Petrie, Assistant Chief Statistician, and John Coombs, Director, Institutions and Social Statistics Branch, about linkage with the 1981 census in particular;
- Ken Bennett, Manager of the Canadian Labour Force Survey, and his colleagues concerning use of the Canadian Labour Force Survey;
- Gustav Goldman, Manager of Planning and Content for the 1991 Census, regarding potential use of the 1991 census;
- Adele Furrie, concerning the Health and Activity Limitation Survey (HALS); and
- John McVey and colleagues about the potential of Combined Records of Employment.

Also, Mr. Neil Barclay of Revenue Canada spent a morning with us discussing income tax records.

Outside of the Federal Government, helpful advice came from:

- Dr. Karl Sieber of NIOSH (the U.S. National Institute for Occupational Safety and Health in Cincinatti, Ohio);
- Dr. Pierre Band of the Cancer Control Agency of British Columbia;
- Dr. Gerry Hill and Ms Shirley Fincham of the Alberta Cancer Board;
- Dr. Marie Swanson of the Michigan Cancer Foundation;

- Drs. Peter Pelmear and Jim Stopps and their staff at the Ontario Ministry of Labour; and
- Dr. Jack Siemiatycki of the Institut Armand Frappier, near Montreal.

As noted in the Introduction, the scope of the review was limited to the creation of databases through linkage of OCR cancer records either with existing occupational data or new occupational data collected on cancer patients. The OCR's capabilities for computerized record linkage (CRL) were examined in some detail. CRL methods in Canada were also reviewed in some depth as additional important background information.

A systematic search to identify sources of occupational data which could be considered for linkage with the OCR was undertaken. This was done through search of government documents and through information obtained at a symposium sponsored by Statistics Canada and held in Ottawa in November, 1987 on "Statistical Uses of Administrative Databases". Appropriate contacts were identified either from information provided with the data source or at the Symposium, or from the government directory. Communication with these contacts by letter and phone led to a trip to Ottawa (on March 24-25, 1988) to further discuss a number of federal databases.

In addition to furthering understanding of the strengths and limitations of the various databases, these meetings resulted in a letter from Mr. Petrie indicating willingness on the part of Statistics Canada to assist in the design of a feasibility study to link the 1981 census with the OCR (the letter is shown in Appendix 3).

Communication was pursued with other individuals responsible for non-federal databases. Each identified source of occupational data was then reviewed with respect to the following:

1. Accessibility:
2. Feasibility of use:
3. Coverage of the working population:
4. Measures of occupational exposures:
5. State of data preparation:
6. Confounder information available:
7. Costs: and
8. What would have to be done to use this source.

The authors then reviewed methods of collecting new occupational data on cancer patients (and, possibly, population controls). Several jurisdictions in Canada and the United States have past or present programmes for the routine gathering of such information. They employ different methods of data collection. Each was reviewed in depth with the evaluation focussing on quality, utility, and cost. Considerations included:

- Validity and reliability of both occupational and confounder information:
- Rates of non-response, including the need for proxies, and the possible resultant biases:
- Methods employed to determine exposures, if any:
- Considerations of statistical power: and
- Costs of collecting and processing data.

Finally, the strengths and limitations of the various methods of generating new data were discussed and an outline for possible approaches within Ontario, with cost estimates, drawn up.

The addition of exposure data to occupational data through the use of a job-exposure matrix is an attractive possibility and could be done for either already collected or newly gathered occupational data. Therefore, considerable time was spent reviewing available matrices, with a view to perhaps using one in Ontario. This led directly into a review of occupation/industry code schemes, which comprised part of the terms of reference of this project. There

was time only for a fairly cursory look at the various relevant systems, as a comprehensive review was beyond the time/fiscal constraints of this study.

For completeness, a few other possible methods of establishing an occupation-cancer database in Ontario were considered very briefly. These included: the use of death certificates; employee records of exposure; and other limited (or future) surveys, particularly the Health and Activity Limitation Survey and the Ontario Health Status Survey.

After all material had been gathered, summarized, and evaluated, a set of recommendations were drawn up, based on the findings. Specific recommendations were made within each major option (on the questions of record linkage, collection of new data, etc.). Broader recommendations spanning options were made for more general issues (viz. the use of job-exposure matrices and occupation coding). Lastly, overall recommendations were developed.

In the Summary of Findings below, each finding and recommendation is presented in summary form. Details may be found in the complete report.

3. SUMMARY OF FINDINGS

3.1 The Ontario Cancer Registry (OCR)

The OCR is a population-based register of all newly diagnosed cancers and all cancer deaths in Ontario residents. It is operated by the Ontario Cancer Treatment and Research Foundation (OCTRF), a statutory foundation of the Ontario Ministry of Health, and is supported by the Cancer Act of Ontario and its revisions (see Appendix 4), which:

- gives the OCTRF the legal mandate to collect such information (Section 5(f) of the Act);
- provides for the maintenance of confidentiality of such information, while permitting use for the purpose of generating statistics and "carrying out medical or epidemiological research" (Section 7(1) of the Act); and
- provides legal protection to organizations or individuals in the health care system who report such information to the OCTRF (Section 7(2) of the Act).

Cancer diagnoses are voluntarily reported to the OCTRF, as there is no legislation requiring such reporting. The OCR includes diagnoses made since 1964. The most recent year for which incidence data are currently available is 1986. The OCR includes information on all cancer deaths occurring in the province among provincial residents between 1950 and 1987. The primary purpose of the OCR is to produce timely, high quality information describing all cases of cancer diagnosed among all Ontario residents.

The methods for generating incidence data for Ontario are presented in detail in Appendix 5. A brief summary follows emphasising those features most relevant to this review.

Presently, the OCR relies on 4 major sources of cancer information to generate cancer incidence reports:

- hospital discharges with one of the 16 permissible diagnoses being cancer;

- pathology reports with a mention of a possible cancer diagnosis;
- death certificates with cancer as the coded underlying cause of death; and
- regional cancer centre reports for patients seen.

All information excepting the pathology reports is received by the OCR in computerized form. Pathology reports are coded and computerized by OCR staff. The four computerized files are then computer-linked using the Generalized Iterative Record Linkage System (GIRLS) developed at Statistics Canada to bring together all reports for the same individual. Two computerized "expert systems" then review all the reports for each individual to select the best patient-specific data (i.e. name, sex, date of birth, etc.) and to determine diagnosis-related information (primary site, histology, date of diagnosis, etc).

Historically, other data sources have been employed either instead of or in addition to the above-described sources. Details of the historical development of the OCR and the processes used to create it may be found in the recently published monograph "Twenty Years of Cancer Incidence in Ontario" (Appendix 5).

OCR computerized incidence data are currently about 2 years behind (i.e. by the end of 1988, data for 1986 were ready), partly due to the dependence on external data sources. About 30,000 new cases of cancer are recorded in the OCR each year.

Identifying information contained in the OCR is obviously dependent on the data source (or sources) in use at any point of time and the number and type of reports present for a given patient. Thus, a patient who was seen recently at one of the regional cancer centres is likely to have the most complete and accurate identifying data, while those reported by pathology report only or by any sources in the early years of the OCR will have the poorest quality data. The

maximal set of identifiers is likely to be for those reported while alive by a reliable source, such as a regional cancer centre, and who have subsequently died (since the Social Insurance Number [SIN], parent names, and birthplace are usually on a death certificate, but are rarely on any other type of record). The most usual complement of identifiers would be surname and given names (and alternates, if known), sex, date of birth, county of residence at various points of time (e.g. for each hospital admission), date and place of death (if deceased), and OHIP number.

The OCR is believed to be very complete for most cancer sites, at least for the more recent years. A special study estimated that over 95% of 1982 diagnoses were reported to the OCR, although this varied by site.

3.2 Computerized Linkage of Administrative and Health Databases

Statistics Canada has a long history of expertise in the area of computerized record linkage, in the acquisition or establishment of potentially linkable databases (e.g. the Canadian Mortality Database), and in the recognition of the utility of linking databases for health purposes. However, it has historically been very cautious in the use of record linkage because of the potential for harm either to individuals (through possible breaches in confidentiality) or to the reputation of Statistics Canada, which could in turn affect compliance with the census and other official surveys. Statistics Canada has a formal statement of policy, outlining conditions under which record linkage activities will be considered and procedures for considering linkage requests. The passage of the Privacy Act and the appointment of a Privacy Commission have made the justification of the need for record linkage more difficult. Thus, Statistics Canada may demand maximal control over linked data and its release to investigators to guarantee appropriate management of identifying information. Basically, it would appear that staff of Statistics Canada are trying to balance the need for research against the need for

privacy, within the current legislation and level of public concern.

The Ontario Cancer Registry also has expertise in computerized record linkage using the Statistics Canada GIRLS. Thus, linkage could be conducted within the OCR provided the OCR could secure access to an appropriate occupational database.

Linkage experts indicate that name (surname and full given names) and sex plus complete date of birth constitute the minimal set of identifiers for linking two files. Other information will also be needed in many cases to confirm or reject possible links. One of the difficulties with linking administrative and health databases is the lack of identifiers in common. While many (but certainly not all) administrative databases include the SIN number, the OCR does not; and while the OCR includes the OHIP number, most administrative databases do not. If the identifiers on the administrative and health databases relate to different points in time (as they would, for example, if 1981 census data were to be linked with subsequent cancer diagnoses), then time-dependent identifiers such as place of residence are of limited utility. The issue of lack of common identifiers on health and administrative databases is a common theme throughout the discussion of various potential occupational databases.

3.3 Potentially Linkable Occupational Databases

Only a few databases were considered in depth as linkable with the OCR and of sufficient quality for the purposes of the Industrial Disease Standards Panel. These are:

- Census of Canada, Long Form (1971, 1981, and 1991);
- Unemployment Insurance (UIC) Book Renewal Cards (based on a 10% sample of the labour force);
- Income tax forms (T1 general and T4);
- Canadian Labour Force Survey; and
- Records of Employment.

Some of the basic characteristics of these databases are summarized in Table 1. Details and in-depth assessment of each may be found in the complete report (Chapter IV, Section 2). There is great variability in the quality of both their occupational and identifying information and in the population coverage achieved. A brief description of each follows.

a) The Census

Three Censuses of Canada were initially considered: those of 1971, 1981, and 1991. Since the 1971 Census did not collect complete date of birth information, it would not be feasible to link it with subsequent OCR data due to lack of common identifiers. Consequently, it is not considered further here.

Twenty per cent of the population received the Long Form version of the 1981 Census of Canada. Names, addresses, and complete dates of birth were recorded, although only postal codes and month and year of birth have been computerized. This long form collected additional information not on the regular census form (viz. info on the work of all household members during the week prior to June 3, 1981). Questions relate to the name and address of the employer: the type of industry; the kind of work done; and the description of main duties. Responses have been coded according to the Standard Occupation Classification and the 1980 Standard Industry Classification for occupation and industry respectively. In addition, all data requested on the regular census form (such as income, education, ethnic origin, place of birth, year of immigration, number of children, year of first marriage, etc), are also included in the long form.

As mentioned, name and day of birth are not computerized so would have to be abstracted from the original microfilmed census records, keyed, and linked with the computerized census files before linkage with the OCR could be carried out. Once such identifiers were

available on the census file, linkage could in theory proceed. The costs of retrieving these additional identifiers and linking the resultant file with the OCR were estimated by Statistics Canada (who would have to do all this work, due to concerns regarding confidentiality) to be about $865,000.

As indicated above, the set of common identifiers between the census and the OCR is minimal and the number of definite links may be too few to be useful. In addition, it may not be possible to resolve a great number of the "possible" links. For this reason, Statistics Canada has recommended that a pilot study be undertaken, designed by OCTRF and Statistics Canada staff jointly, to determine whether the proposed linkage would indeed be feasible and useful, and whether it could be done in such a way as to conform with Statistics Canada's requirements regarding confidentiality while still yielding data useful to the OCTRF and the IDSP. One of the likely constraints to be imposed by Statistics Canada is the provision of tabulated data only, rather than individual linked records.

Provided that identifying data are of sufficient quality and quantity that good links between census and OCR files for the period 1981-1989 can be identified, the expected number of cancers among Ontario males aged 20-64 at the time of the census (about 0.5 million) would be between 9,500 and 13,250, depending on whether a 10% or 5% annual loss due to death and/or emigration from Ontario is assumed (See table IV.2 in complete report). If the larger of these figures is taken, 350 or more links would be available for each site constituting more than 2% of the incident cancers in men, including cancers of the lung, prostate, colon, bladder, etc. (See table IV.3 in complete report). In fact, if the methodology works, the study could easily (but at some cost) be expanded to cover the entire country. Epidemiologists in other provincial cancer registries have expressed cautious interest. (Concern about costs is responsible for the caution.) Of course,

the power would increase with increasing years of follow-up also. Estimates of power for detecting associations between various occupational groups and specified cancer sites are given in the complete report.

Consideration was also given to use of the 1991 census, because it was believed that prospective data entry of additional identifying information may be more efficient than retrospective retrieval and keying. Presently collection and keying of identifiers is planned to be similar to the 1981 census, with only month and year of birth keyed, and name not keyed at all. Use of the 1991 census was not pursued further since:

- Keying of name would increase entry costs substantially (and would not be borne by Statistics Canada);
- Statistics Canada has indicated categorically that there will be no linkage between censuses (ie 1981 and 1991 censuses, for example);
- Changes to content of 1991 census can no longer be made (the deadline was sometime in 1988); and
- Use of 1991 census data would necessitate waiting until well into the next decade before enough cancer cases would link with census.

b) Unemployment Insurance (UIC) Book Renewal Cards

A computerized data file exists of a sample of the labour force for the period 1965-1971. This is called the UIC Book Renewal Card file. Data were provided annually by employers on all employees whose SIN number ended with the digit "4" during the period 1965-69 and for all those whose SIN ended with a "4" preceded by an odd number in 1970-1971, when social insurance was expanded to include white collar workers. The primary identifiers on this file were originally SIN and name. Through a linkage with the SIN master file, other identifying information was added, such as date of birth and mother's maiden surname. Occupation data are provided in an open-ended fashion by the employer and are subsequently coded.

Industry comprises a description of the nature of the business and is also coded subsequently.

Dr. Geoff Howe of the National Cancer Institute of Canada's Epidemiology Unit (at the University of Toronto) and Ms. Joan Lindsay of Statistics Canada have linked the 1965-69 portion of the file with the Canadian Mortality Database. Four-year (or less) job histories have been created by linking records for the same individual over time. It is likely that they have plans to link this file with the National Cancer Incidence Reporting System (NCIRS), which will contain computerized identifying and cancer information from all the provincial/territorial cancer registries in Canada, when it is fully operational. Note that for the years 1965-69 this is a 10% sample of the labour force This represents quite a small sample for the province of Ontario alone.

c) Income Tax Records

A personal income tax return is filed with Revenue Canada annually by every person in Canada who has income to declare. There are two forms that may be used: the T1 General, and the T1 Special. The T1 Special is a short form for filers whose tax affairs are very simple and straightforward, with only usual deductions or income sources. It is used by about 1/3 of filers and, due to its brevity, does not include information on occupation. The T1 General is completed by about 2/3 of filers and must be used by those whose deductions and income sources are more complicated. This form requests occupation ("Type of work or occupation in tax year") and employer name. There are no guidelines to the completion of this information, and no quality control. The information is not entered on an estimated 25% of the T1 Generals. The information is used by Revenue Canada officials only to determine whether declared income is reasonable.

Revenue Canada also administers the T4 slip which is completed by the employer for each employee. Each slip contains the employer's

"PD" number which can be linked to the Business Registry to retrieve the Standardized Industrial Classification code of the primary industry of the employer. Thus an industry code is available for each employed person.

Selected data are keyed from T1 forms, including SIN as the primary identifier. Occupation/industry data from T1s are not computerized. In order to conduct linkage with these it would be necessary to carry out an intermediary linkage with the SIN master file to add identifiers. Then income tax forms would have to be retrieved for OCR links and occupation information manually abstracted and keyed. Paper forms are stored for 4 years only.

It has been concluded that these income tax records do not presently suit the purpose because of:

- The dubious quality of income tax occupation data;
- The large number of forms with no occupation information, either because a T1 Special was used or because no information was provided;
- The likely bias associated with missing information in the case of filers using the T1 Special;
- The limited time into the past that information is retained; and
- The general reluctance, lack of interest, and confidentiality concerns of Revenue Canada staff.

Only if Revenue Canada could be persuaded to collect better and more complete occupation information and to key at least the alphameric occupation data from the tax forms along with tax-related information would this data source be worth pursuing at some future date. Presumably, employer information provided on T1s is redundant with T4 information and could be dropped. The wide coverage is the main appeal of this data source.

It is worth noting that major efforts have been and are being made in the U.S. to evaluate the quality of occupation data provided on

tax forms and to investigate the feasibility and utility of using these data in health studies.

d) Canadian Labour Force Survey (CLFS)

The CLFS is a nationwide, continuous household survey whose primary purpose is to provide monthly figures on the size and composition of the labour force. Occupation information is collected every month for a period of 6 months for all members of selected households who are over the age of 15 years. Questions concern work in the week prior to the survey and include a description of work done and main job activities, as well as name of employer and description of industry and start date of current job. These data are coded according to the Canadian Standard Occupation Code (SOC) and Standard Industrial Classification (SIC) systems.

The sample covers about 1 in 400 or 500 households in Ontario each cycle (6 months). This translates into less than 5% of the adult population over a period of 10 years. Information is often provided by a proxy for the entire household. Supplementary surveys are regularly added on to the CLFS. For example, smoking data are collected every 2 years on one third of the LFS sample. Limited identifiers are collected and keyed, namely, name (surname, given name), age, sex, and address. Unfortunately, complete date of birth is not included. Thus, linkage would be essentially impossible unless complete date of birth were to be collected. CLFS staff indicated a willingness to consider adding date of birth, but the number of individuals included in the survey is really quite small.

e) Combined Employment Records

Occupational and industrial information may be generated by combining several databases maintained by Canada Employment and Immigration (CEI) and Revenue Canada. These are collected on specifically defined populations primarily for administrative

purposes and therefore do not cover a random sample of the work force.

The basis of such a database would be the Records of Employment (ROE). An ROE is filed with CEI by an employer each time an employee separates (i.e. terminates his/her job with that employer), basically for the purpose of administering unemployment insurance benefits. Occupation is recorded on the form, using an open-ended question ("Employee's occupation"), as are SIN, duration of employment in the job, name and address of employee, and name, address and PD number of employer. Although ROEs are computerized, occupation and name and address of employee are not included. ROEs have been microfilmed for at least 10 years. Staff of Statistics Canada are currently abstracting occupation information from a sample of ROEs from the microfilm and are trying to code them. This will result in an evaluation both of the quality of occupational data on the forms and the ease of coding, although no assessment of validity or extent of coverage will be possible.

Statistics Canada are trying to improve the quality of these data and would like to see occupation data keyed routinely. They would also like to see linkage/merging with other databases to improve coverage and quality of data. Suggested databases include: income tax records (T1s) for self-employed individuals: information on hirings (provided to CEI by a sample of employers on a pilot basis: coverage could be expanded to all employers): and T4s. The advantage of these databases would be that occupational histories could be constructed.

Linkage with external databases would require intermediary linkage with SIN to add name and date of birth.

3.4 Generation of New Occupation Data for Cancer Patients

Collection of occupational histories from newly diagnosed cancer patients reported to the Ontario Cancer Registry was considered

through use of mail questionnaires, telephone interviews, or in-person interviews of subjects themselves or their proxies. A summary comparison of four geographic regions using these approaches is found in Table 2. Details may be found in Chapter V of the larger report. A major advantage of collecting new data is the possibility of acquiring information on potentially confounding variables (e.g., smoking, drinking, ethnicity, etc) as well as job histories and more details about job duties and, possibly, workplace exposures. The major disadvantage is the cost involved in gathering and coding such data.

The Cancer Control Agency of British Columbia and the Alberta Cancer Board have both been employing mailed questionnaires to newly diagnosed cancer patients reported to provincial registries to collect occupational and confounder information, while the Michigan Cancer Foundation uses telephone interviews with patients in the Detroit Cancer Registry. Both methods seem to achieve good results in terms of response rates and quality of data, although the Detroit operation claims to have a much lower cost per case compared to B.C. and Alberta. It is, however, difficult to be sure of what different operations are including in their cost estimates. The best estimate is that the cost would not be substantially different for the two methods of data collection in Ontario, and are comparable to costs being incurred in B.C. and Alberta (namely about $75 per case).

Dr. Jack Siemiatycki has, on the other hand, been collecting occupational information from cancer patients in the Montreal area through use of in-person interviews (usually conducted in hospital). He collects detailed occupation and exposure data for selected exposures and has industrial hygienists/chemists review each history and estimate exposures (and their intensity) for each person-job. Needless to say, this is very costly. However, Dr. Siemiatycki has demonstrated the increased power achieved by his exposure estimation when analysis is done by exposure

classification. No validation of his exposure estimates is available, although inter-rater variation has been assessed. From this study, a job-exposure matrix could be constructed which would be appropriate for industries in the Montreal area over the past 40 or so years. Dr. Siemiatycki has suggested that his staff could assess the applicability of this matrix for the province of Ontario.

Experience conducting studies including occupational histories in Ontario suggests that it would be feasible to collect such information from cancer cases on a routine basis. The OCR can be used to identify eligible cases. Physicians have been cooperative (their consent to contact their patients is required), and cases have responded at relatively high rates. Both mail and interviewer-administered in-person questionnaires have been utilized extensively by OCTRF researchers. They have less experience conducting telephone interviews but are beginning some studies employing this method now and have no reason to expect lower levels of cooperation or data quality. In a large province like Ontario, substantial numbers of cancer cases can be accumulated over a short period of time for many cancer sites (e.g. over 300 males aged 25-74 are diagnosed with the following cancers each year: stomach; colon; rectum; pancreas; larynx; lung; melanoma; bladder; kidney; non-Hodgkin's lymphoma; and leukemia).

3.5 Enhancement of Occupational Data

All the considered methods involve the use of data on occupation. As indicated earlier, it is really exposures occurring on the job that are of concern, not the occupation itself. However, it is easier to consider using occupation as a surrogate for exposure for several reasons:

a) Data on occupation are already collected by census and other survey groups, whereas exposure data are rarely available:

b) Individuals can probably more accurately report

occupation than many exposures, especially those that were in the distant past:

c) It is much cheaper to collect occupation than exposure data:

d) Occupational data permit a variety of hypotheses to be tested, not only on exposures, but also on socio-economic class, ergonomic factors, working conditions, etc.

Over the past decade or so "job-exposure" matrices have been developed to improve the specificity of exposure information available from occupational data. Basically, these matrices provide a series of "links" between job titles (usually coded) and on-the-job exposures. These may be developed by reference to the literature: by consultation with chemical engineers and industrial hygienists: by actual observation of work place operations: or by some combination of these. Siemiatycki has demonstrated the increased power for examining exposure-disease associations which can result from use of a job-exposure matrix imposed upon an occupational history, in comparison with use of the job titles alone.

Several job-exposure matrices were explored for this report. The one developed by staff of the National Institute for Occupational Safety and Health in the U.S.A. (NIOSH) appears to be best for use in Ontario. A copy of this matrix has been acquired (at no cost) by the investigator, although further investigation is needed to determine how useful it would really be for Ontario. For example, occupations are coded according to the US Census occupation coding system, which may not be compatible with Canadian codes: the matrix is based on US industry and may not be applicable to Ontario for all industries: the matrix was constructed according to exposures occurring at the time of its creation and these may not be applicable to the same job titles many years ago: minimal "field evaluation" of the matrix has occurred.

The other possibility for a job-exposure matrix stems from the work of Dr. Siemiatycki. From his study databank he could create a job-exposure matrix which he believes to accurately reflect the experience of Montreal residents over their working lifetime. Some changes/additions would undoubtedly need to be made to make it applicable to the Ontario population. In addition, Dr. Siemiatycki uses a slightly different occupation coding system than that used by the census which would require some conversion. He has estimated that it would cost about $20,000 for his group to produce a comprehensive job-exposure matrix from his databank and to expand/modify it to fit the Ontario industrial scene.

3.6 Exposure Databases

While recognizing that the ideal occupation-cancer database would include exposure information, or would actually be a workplace exposure-cancer database, it is extremely difficult to get accurate workplace exposure data on a representative sample of the Ontario population. Such information is available for subgroups of the population with specific exposures (e.g. atomic workers including uranium miners, etc.), but even for monitored groups such as these it has not been easy to assemble exposure data.

Some regions (e.g. Finland) have attempted to develop "exposure registers" for specified exposures. It has been proposed that such a system could be developed for Ontario, based on legislation related to compliance with the recent worker/community right to know amendment to the Ontario Occupational Health and Safety Act. The Workplace Hazardous Materials Information System (WHMIS) legislation (O.Reg. 644/88) constitutes the first such regulation, and sets out part of a national program which is intended to provide workers and employers with information about the hazardous substances with which they work for the purpose of protecting health and safety. Its main thrust is education which is carried out by way of three activities: warning labels on containers,

material safety data sheets (MSDS) on each hazardous material, and worker training on how to use the information.

MSDS sheets contain information on the chemical, its formula, handling and transport, toxicity, emergency treatment, etc. These MSDS are to be readily available or posted in designated locations. Labels on the container are to be of standard design indicating handling, transport, toxicity and emergency treatment. Regular education programs are to complement this information.

Each workplace will keep a list of all such materials and information. These lists are accessible by the local public health agency as the intermediary between the workplace and the community. Community right to know, therefore, can be implemented by access to the information by the local medical officer of health.

WHMIS is a system which can theoretically enable information to be available on potential or measured workplace exposures for every worker over time. The presence of lists in each workplace can theoretically provide the basis for exposure potential in the study of health effects in a given work population.

However, there is no plan at the moment to implement a central information system whereby employers can provide a continuous listing of materials handled within each workplace. A continuously updated list would be needed as a particular work place might change the materials which it handles over time. Anything short of a centralized, computerized system which would allow such records to be stored would not be useful. Such would be needed for the stratification of workers by a closer measure of chemical exposure for the analysis of risk rather than by more distant surrogates of exposure such as type of job or industry possibly augmented by a job exposure matrix. While it could be made mandatory for all employers to file copies of their exposure datasheets (with updates) with a central repository, enforcement could be difficult

and compliance is unlikely to be high without subsidies and assistance. Standardization of the data could be time-consuming.

There is currently nothing in the legislation which requires employers to retain or forward information on exposures of any of their employees by name. In order to have an exposure-employee database, companies would have to be required to retain (and forward to a central agency) such information in a standardized way with enough identifiers to permit subsequent linkage. While this could be made obligatory (and compliance is unlikely otherwise), incentives, subsidies, and assistance with standardization and computerization would probably be necessary to increase compliance. A number of questions arise when considering this option:

- Should this be selective with respect to exposures included?
- Should historical information be included?
- How would records be updated?
- Who would be responsible for all this, including the cost?

Since legislation is not currently in place for such an exposure register and since development of such a register would require tremendous cooperation on the part of many employers, even for selected exposures, it was felt that further development of this option was beyond the scope of the present study.

4. RECOMMENDATIONS

The recommendations flowing from this review follow. The first (Recommendation A, the exploration of the feasibility of linking census and OCR data to form an occupational cancer data base) is clearly the authors' preferred choice for the next phase of this project, although Recommendation C (the development of an exposure register) should probably be followed up concurrently as it is really addressing present hazards rather than past ones. Rec. B (the feasibility of linking other Statistics Canada data files with the OCR) and Rec. D (the feasibility of collecting new data from cancer cases) are possible alternatives to Rec. A, should it prove not to be feasible or satisfactory. Probably both of these should be addressed simultaneously, since it is not clear that Rec. B would result in anything tangible.

RECOMMENDATION A:

The development of a pilot study to examine the feasibility and utility of linking census and OCR data should be pursued with staff of Statistics Canada. As part of this, the possibility of including other provincial cancer registries in linkage of census and cancer incidence data should be explored. Fairly precise estimates of the cost of carrying out such a linkage should be developed as part of the pilot study.

Of all the occupational databases explored, the 1981 census is considered to have the most potential utility for establishment of an occupational cancer database for the province. Reasons for this conclusion include:

- Coverage (20% of households, selected more or less randomly);
- Quality of occupation data (contains specific questions which are fairly detailed, and a high rate of response);
- High quality coding (done by well-trained staff at Statistics Canada);
- Availability of some demographic data for confounder control (e.g. ethnicity, socio-economic status, etc.).

Provided adequate identifiers can be obtained, the expected number of links with the OCR would be large enough by 1989 to study associations between a number of cancer sites and occupations. Continued follow-up through the 1990s would serve to increase power, as would use of a job-exposure matrix to decrease the amount of exposure misclassification (See also Recommendation E).

Limitations of using census data include the paucity of identifiers in common with the OCR, making linkage potentially difficult; lack of information on important lifestyle confounders such as smoking; no job histories (only occupation at time of census); occupational exposure information limited to that contained within occupation/industry data provided; some proxy responses (since one form includes all members of the selected household).

At present, insufficient census identifiers are computerized to permit any linkage attempt. Therefore, before proceeding with linkage, name and complete date of birth (only month and year have been computerized) would have to be abstracted manually from census forms, computerized, and linked with computerized census data.

Up until this time, linkage of census and outside databases has not been considered due to concerns regarding privacy and confidentiality. However, as indicated in the letter from Mr. Bruce Petrie, Assistant Chief Statistician, Statistics Canada (see Appendix 3), "...Statistics Canada is prepared to cooperate with the OCTRF in an effort to identify how such a database [Editor: that is, one created through linkage of 1981 census data and subsequent cancer incidence] might be established in a way which would both meet your [viz. the Panel's] research objectives and respect Statistics Canada's concerns and requirements with respect to matters of confidentiality and privacy.". This should be considered a major breakthrough in terms of access to data, although enthusiasm must be restrained because of the approvals

still required for an actual linkage and the level of control which Statistics Canada must be permitted to retain (e.g. microrecords would not be released, but only data tabulations). A pilot study would need to work out the details regarding:

- Feasibility of the linkage itself (due to possible lack of identifiers);
- The utility of the approach, in terms of quantity and quality of resultant data;
- The acceptability, in terms of the degree of control which Statistics Canada will wish to exercise; and
- The cost.

The cost of designing a pilot study would be about $6,000. This figure includes: investigator time (estimated at 10 days); clerical assistance (i.e. typing correspondence, coordinating meetings, and researching details as necessary); telephone; travel (involving 3 trips to Ottawa); and general office supplies and costs. Detailed discussions with Statistics Canada staff would be needed to estimate precisely the time/resources required for the pilot study.

Other provincial cancer registries have indicated a willingness to participate in a census-cancer registry linkage should this initiative take place, contingent as always on final cost estimates. The authors believe this is a worthwhile pursuit.

The pilot study should include consideration of ways of evaluating the potential limitations of census data. For example, the feasibility of linking a sample of census forms with the UIC Book Renewal Cards file could be explored. This would permit assessment of census occupation as an indicator of prior occupational exposures. In addition, the possibility of using Canadian Labour Force Survey smoking data by occupation in conjunction with linked census-OCR data to assess the potential role of confounding by smoking should be explored. Finally, the feasibility of assessing

the validity of reported industry by linkage of a sample of census forms with T4s could be examined.

RECOMMENDATION B:

The possibility of linking other files containing occupation data (viz. the Canadian Labour Force Survey, income tax data, or Records of Employment) should be explored further with Statistics Canada/Revenue Canada staff.

As indicated above, all these files currently have drawbacks in terms of their use as the primary source of occupation information in the establishment of a cancer-occupation database for Ontario. However, it is possible that future changes could make one of them a better source than the census, particularly if linkage with the census turns out to be infeasible or unsatisfactory for some other reason. What requires further exploration, therefore, are:

- The changes which would have to be implemented to make these databases acceptable sources for linkage (viz. improvement of identifiers on the CLFS);
- The improvement of questions and key entry of occupation data for income tax data;
- Increased coverage of COE; and
- Issues related to accessibility (particularly for tax data).

Income tax has the most potential (wide coverage, good identifiers, potential for creation of job histories) but the most problems also (related to access, rationale in terms of Revenue Canada's mandate, and cost).

RECOMMENDATION C:

The development of a comprehensive exposure register which includes good identifying information should be encouraged for future surveillance and future development of an empirical job-exposure matrix for Ontario.

As indicated above, there is no easy way to systematically access exposure information on a random sample of Ontario workers. However, with the new right-to-know legislation, the political climate and worker concern may be sufficient to move to develop a comprehensive exposure register. Some industrial groups are already included in registers maintained either by the Ministry of Labour or research groups interested in occupational health. Expansion could be gradual to include other hazardous exposures a few at a time, as in a feasibility study. Issues related to industry compliance and the bearing of the continuing costs of establishing and updating such a register will have to be considered. Standardized reporting must be developed and the minimal set of identifying information must be determined.

RECOMMENDATION D:

A pilot study should be conducted to determine more precisely the cost and quality of new occupation (and confounder) data collected by mail versus telephone.

While it is clear that data could be collected either by mail or by telephone in Ontario, the resultant quality of data and costs of collection are not precisely known. If at some future time the Panel decided to proceed with the collection of new data, a pilot study would be needed to develop the study instrument, to test alternative methods of administration (telephone versus mail), and to obtain more precise estimates of the cost of collecting data on a very large number of people. Issues related to use of a population control group would need to be resolved also, and sources/sampling strategies explored.

This method doubtless results in useable data for the creation of a cancer-occupation database. The best estimate of its cost is comparable to that for census-OCR linkage. The advantages and disadvantages of collecting new data versus linking census data to the OCR are discussed more thoroughly in the main report (chapter

IX), but may be briefly summarized here. The main advantages of the census are that it has an extremely high response rate, and data are generally provided by the individual his/her self (rather than a proxy, as is often the case in collecting new data). Occupation data are collected currently (thus eliminating any recall problems) and before the diagnosis of cancer (no bias). Linkages between the census and the OCR can continue for many years into the future at minimal cost, thus increasing the power of studies based on these data. On the other hand, advantages of newly collected data are that occupational histories can be constructed, resulting in the assessment of lifetime exposures. Information on important confounders can also be collected.

RECOMMENDATION E:

The suitability and adaptability (for Ontario) of available (or easily created) job-exposure matrices should be further explored, to increase the power and specificity of any occupation-cancer database. This will necessitate investigation of conversion of various occupation coding systems. Use of automated occupation coding should be further explored also.

As indicated above, job-exposure matrices (JEM) can substantially increase the power and specificity of occupation-cancer associations. The JEM developed by NIOSH has been secured and an investigation of its utility in conjunction with an occupational case-control study is being planned. Further exploration of Dr. Siemiatycki's JEM could also be made.

Since each dataset seems to employ a different classification system for occupation, investigation of coding conversions is a necessary part of this endeavour. Plans are already underway to start this for the NIOSH JEM.

Automated occupation coding would facilitate the handling of newly collected occupation data and would improve the consistency of

coded data (i.e. ensuring that the same occupation would always receive the same code). Statistics Canada has developed such a system for coding ROE data: it may be adaptable for coding study data. It would also be useful for coding occupation data on sources currently collecting but not coding such data, such as income tax forms. If an automated coding system were in place, Revenue Canada staff would only need to key occupation as reported for subsequent computer coding. No automated system can code every entry, but most aim for at least 60-80%.

RECOMMENDATION F:

A unique, personal, lifetime identification number should be developed for Ontario residents and used within both the health and administrative spheres to facilitate linkage.

One of the problems with linkage of administrative and health files is the lack of identifiers in common, especially if they relate to events at different points of time (so that residence, for example, is not a useful piece of identifying information). Linkage would be greatly facilitated (and with a much higher "hit" rate) if there were one personal number (as opposed to several - SIN, OHIP, etc.) assigned to each individual for life (as opposed to changing with various changes of status). This recommendation is not one that can be acted upon, except by lobbying the political system within Ontario for change.

TABLE 1

DESCRIPTION OF POTENTIALLY LINKABLE OCCUPATIONAL DATABASES

	Proportion of Ontario	Years	-Current- Occupation	-Current- Industry	----------Personal Identifiers---------- Personal available	Identifiers automated?
1991 Census Long Form	1/5 general population	1991 (snapshot)	yes - coded	yes - coded	name D.O.B. - date, month, year postal code	no maybe yes yes yes
1981 Census Long Form	1/5 general population (random)	1981 (snapshot)	4 digit SOC	4 digit SIC	name, D.O.B. - date, month, year sex postal code	no no yes yes yes yes
1971 Census Long Form	1/3 general population (random)	1971 (snapshot)	4 digit 1970 OCM	4 digit SIC	name D.O.B. - date, month, year sex postal code	no no no yes yes yes
UIC Cards	10% UIC Contributors (700,000 work histories Canada-wide)	longitudinal sample 1965-69	3 digit 1961 OCM	1960 SIC	SIN	require inter-mediary link with SIN Central Index
T1 General	2/3 of individuals with income to declare	retained for 4 years by Revenue Canada	yes - not captured or coded	not on T1	SIN	require inter-mediary link with SIN Central Index

TABLE 1 (CONT'D.)

DESCRIPTION OF POTENTIALLY LINKABLE OCCUPATIONAL DATABASES

	Proportion of Ontario	Years	-Current- Occupation	-Current- Industry	----------Personal Identifiers---------- Personal available	Identifiers automated?
T4	all employed persons	1972-	no	SIC	SIN	require intermediary link with SIN Central Index
CLFS (past)	1/400 - 1/500 5% over 10 years	rotating sample	75-82 2 SOC 84-85 3 SOC 86 4 SOC	3 digit	name age postal code	yes yes yes
CLFS (future)	"	"	4 digit SOC	3 digit SIC	name D.O.B.*	yes possible
RECORD OF EMPLOYMENT	6 million/year (Canada-wide) (Nos.of persons?)	1978-›onwards	4 digit SOC	(indirectly)	SIN	link with SIN Central Index

NOTE: * possible

TABLE 2

COMPARISON OF POPULATION-BASED STUDIES SYSTEMATICALLY COLLECTING OCCUPATION DATA ON NEWLY-DIAGNOSED CANCER PATIENTS

	BRITISH COLUMBIA	ALBERTA	MONTREAL	DETROIT
Date of Inception	January, 1983	July, 1983	1979	November, 1984
Sample: Males	newly diagnosed cases age 20+ years	newly diagnosed cases aged 25-74 years	newly diagnosed cases age 35-70 years	newly diagnosed cases aged 40-84
Females	commencing study	25% random sample of females	---	newly diagnosed cases aged 40-84
Procedure	postal survey	postal survey	personal interview	telephone interview
Approx # of Cases available for Analysis	13,000(1988)	4,218(1987)	3,700(1985)	5,734(Aug. 1986)
Response Rate	63.00%	75.00%	82.00%	92.6%(proxies included)
Date:				
-Job Titles	x	x	x	x
-Description of duties, etc.	x	---	x	x
-Industry	x	x	x	x
-Duration	x	x	x	x
-Exposure	-not self reported -industrial chemist codes exposures for SOC job titles -linked with survey data by job title	-respondent completes exposure checklist	-respondent describes work environment/ processes -team of chemists complete exposure checklist	-none

TABLE 2 (CONT'D.)

COMPARISON OF POPULATION-BASED STUDIES SYSTEMATICALLY COLLECTING OCCUPATION DATA ON NEWLY-DIAGNOSED CANCER PATIENTS

	BRITISH COLUMBIA	ALBERTA	MONTREAL	DETROIT
Confounders:				
-Tobacco	x	x	x	x
-Alcohol	x	x	x	x
-Age (D.O.B.)	x	x	x	x
-Sex	x	x	x	x
-Current resid.	x	x	(probably)	x
-Marital Status	x	x	(probably)	x
-Ethnicity	x		x	x
-Education	x	x	x	x
-Other		birthplace religion residential history employment status family size family medical history	residential history home environment socioeconomic status coffee/tea foods containing carotene height/weight	residential history birthplace health status items
Controls	other cancer sites	other cancer sites	population and other cancer sites	other cancer sites
Costs/case*	$75.00	$75.00	$450.00	$25.00

* These figures are very rough estimates. They do not seem to include starting up costs.

APPENDIX 1.

TERMS OF REFERENCE FOR THIS REVIEW

a) IDSP's Request for Proposal

b) Proposal as Submitted by L.D. Marrett

July 2, 1987

INDUSTRIAL DISEASE STANDARDS PANEL

DEVELOPING AN OCCUPATIONAL CANCER DATA BASE

TERMS OF REFERENCE FOR FEASIBILITY STUDY

1.0 The primary objective of the feasibility study is to determine whether an occupational cancer data base can be developed by linking occupational data with appropriately matched records of the Ontario Cancer Treatment and Research Foundation (OCTRF). Ultimately, the Industrial Disease Standards Panel wants to have produced 'maps' revealing the burden of occupational cancer by S.I.C. industry group or subgroup (to the 3 or 4 digit level if possible) and by a standardized occupational coding scheme. This feasibility study will result in a written report (containing relevant technical appendices and recommendations) laying out alternative approaches or methodologies, and their related costs, benefits, staffing and other resource requirements and time frames, for yielding occupational cancer risk estimates (both morbidity and mortality).

2.0 This report should also address:

2.1 EXPECTED BENEFITS: Anticipated benefits should be defined by drawing upon: existing efforts to employ the OCTRF cancer registry as an occupational

cancer data base; corresponding models in other jurisdictions (viz. the Howe/Lindsay approach through StatsCanada, other provincial or national approaches, American and international approaches, etc.). Expected benefits should be related to alternative methodological approaches and to related costs and time frames (viz. prospective cohort methods, case control studies, population attributable risk studies, etc.). What are the likely indicators of occupational morbidity and mortality to emerge from these approaches?

2.2 METHODOLOGY

Methodologies for obtaining linkable data (through computerized record linkage, interviews, etc.) and the corresponding yield of estimates of risk by disease site which will result should be fully defined and explored in the report. The feasibility of each defined approach and the likely kinds of errors and their estimated magnitudes should be identified. Each approach should identify as well the level of possible industry coding (for the 2,3 or 4 digit SIC level) and the most appropriate standardized occupational coding scheme. The standardized

occupational coding should be related to the primary and secondary manufacturing base of the Ontario economy and to the shift to a more service oriented economy. A survey of the different occupational coding schemes employed in other jurisdictions may bring to light innovative approaches to coding so the report should engage in comparison studies.

2.3 DATA: The report must include information on all data required for each alternative approach. This information must indicate as fully as possible the contents of all the data accessed or created and must comment on any likely problems in gaining access to this data.

2.4 TIME: The report must put forward a time-frame for implementation of each alternative approach. This should include discussion of each overall project (or approach), its logical phases or stages and estimated time duration for each such phase.

2.5 STAFFING: The report must indicate staffing requirements for each proposed project,

including proposed job descriptions for indicated positions and anticipated salary levels.

2.6 COST:

For each defined approach, projected costs should be broken down by project phase, staffing costs, related computer development and maintenance costs and related acquisition costs (capital equipment and/or data requirements).

REVIEW OF METHODS FOR DEVELOPMENT OF AN OCCUPATIONAL CANCER DATA BASE IN ONTARIO

Response to Terms of Reference Document of the Industrial Disease Standards Panel

Prepared by: Loraine D. Marrett, Ph.D.
Senior Epidemiologist
Division of Epidemiology and Statistics
Ontario Cancer Treatment and Research Foundation

and
Director, OCTRF Epidemiology Research Unit
Department of Preventive Medicine and Biostatistics
University of Toronto

Date: September 15, 1987 Revised: October 26, 1987

I. SUMMARY

This proposal addresses the objective outlined in the Terms of Reference Document referred to above, namely the conduct of a review to determine whether an occupational cancer data base can be developed for the province of Ontario which could be used to estimate the burden of occupational cancer within specific occupational and industrial groups. The review which is proposed will outline possible approaches to the quantification and "mapping" (by occupational/ industrial group) of occupational cancer in Ontario. For each approach outlined, the necessary data will be described and their availability explored; the expected advantages and disadvantages will be detailed; and the anticipated requirements in terms of resources (staffing, costs) and time frames will be discussed. It is acknowledged that interest lies in both cancer morbidity and cancer mortality and in both the present and future burdens of cancer in Ontario.

The study will review the various approaches which have been used in other jurisdictions as well as possibilities unique to Ontario, which result from particular circumstances or data systems already in place or readily available.

II. OUTLINE OF PROPOSED FEASIBILITY STUDY

The feasibility study can be divided into 2 main components, which follow each other in a logical progression, namely:

1. Gathering of information: During this phase all the information necessary to making recommendations and evaluation will be gathered.

2. Outline of possible approaches: Based on the information gathered in phase one, possible approaches to the quantification of occupational cancer in Ontario will be outlined.

Each of these will be described more fully below.

1. Gathering of Information

The proposed major methods of gathering the information necessary to outline and evaluate approaches to the quantification of occupational cancer in Ontario are as follows:

a) **Review of methodologies:** This includes both review of the literature and personal communications with selected individuals. There have been previous attempts to quantify or characterize occupational cancer in defined populations. Results of some of these have been published in the literature (e.g. proportional mortality analysis in Washington State), while some are ongoing and have no or minimal information available to the public (e.g. efforts of the Cancer Control Agency of British Columbia to collect occupational histories on selected cases of cancer in the province). It is important not only to review efforts of those known to us, but also to try to find out about work of others in the field. It will thus be necessary to "network", or ask those we do know of about other work going on.

b) **Review of existing databases which may be of use:** There are some databases which contain cancer information and many which include at least some occupational data. Databases known to us will be investigated with regard to their potential utility and additional databases which may exist will be searched for. Quality, suitability, and availability of various databases will be ascertained with respect to present or future use; if present use is not feasible and the database appears to be useful, then requirements to ensure future availability will be outlined.

c) **Review of coding schemes for occupation and industry:** Coding schemes employed by those within the sphere of occupational health and by others interested in collecting statistics (eg census) will be reviewed and evaluated with respect to utility for the problem at hand.

d) **Collection of new ideas:** Those working in the field of occupational health, particularly as it pertains to cancer, will be interviewed regarding their ideas regarding quantification of cancer. Some new ideas may devolve from these interviews.

2. Outline of possible approaches

Based on information gathered in 1. above, possible approaches to the quantification of occupational cancer in Ontario will be outlined. The following will be addressed when describing each approach:

a) **data required** - will include a description of the data sources; a detailed description of the data themselves; access and availability.

b) **coding scheme** - will outline the coding scheme/level recommended for use (or already in use).

c) **methods** - will describe the methods of data collection, linkage, and analysis, including the measures of occupational morbidity/mortality to be used.

d) **time frame** - will provide estimates of the time required to implement each approach, with division of the project into phases and estimation of phase-specific times if appropriate.

e) **staffing requirements** - will estimate staff requirements, including brief job descriptions and recommended levels and types of training.

f) **costs** - will outline approximate anticipated costs, by project phase where appropriate, including salaries, computer costs, and acquisition costs (for equipment and/or data).

APPENDIX 2.

List of Contacts Made During this Review

Appendix 2: Contacts made during this Review

Record Linkage

Mr. John Silins
Chief, Vital and Health Statistics Section
Health Division
Social, Institutions and Labour Statistics Field
Statistics Canada

Ms. Martha Fair
Head, Occupational and Environmental Health Unit
Vital and Health Statistics Section
Health Division
Social, Institutions and Labour Statistics Field
Statistics Canada

Ms. J. Podoluk
Consultant

Census of Canada

Mr. Bruce Petrie
Assistant Chief Statistician
Social, Institutions and Labour Statistics Field
Statistics Canada

Mr. John Coombs
Director, Institutions and Social Statistics Branch
Social, Institutions and Labour Statistics Field
Statistics Canada

Dr. David Bray
Director, Health Division
Institutions and Social Statistics Branch
Social, Institutions and Labour Statistics Field
Statistics Canada

Mr. Gustav Goldman
Manager, 1991 Census of Population
Statistics Canada

Canadian Income Tax Records

F. Hostetter
Statistical Services Division
Treasury Board of Canada
Revenue Canada

Neil Barclay
Revenue Canada Taxation
Revenue Canada

Dr. John Leyes
Director, Small Area and Administrative Data Branch
Statistics Canada

Ms. Joan Berry
Senior Manager
Guaranteed Income and Tax Credit Branch
Ministry of Revenue
Government of Ontario

Mr. Peter Bernard
Manager, Planning and Development
Data Services and Development Branch
Ontario Ministry of Revenue

Canadian Labour Force Survey

Mr. Ken Bennett
Manager, Labour Force Survey Subdivision
Household Surveys Division
Statistics Canada

Mr. Scott Murray
Supplementary Surveys Group
Labour Force Survey Subdivision
Household Surveys Division
Statistics Canada

Ms. Alison Hayle
Labour Force Survey Subdivision
Household Surveys Division
Statistics Canada

Combined Records of Employment

Mr. John Mcvey
Statistics Canada

Ms. Brenda Hutchinson
Statistics Canada

Mr. Benjamin Hazzan
Director, Data Development Division
Employment-Immigration Canada

Other Employment and Immigration:
Dougall Aucoin
David Galliland
Jacques Bourdage

Other Potential Sources of Occupational Data/General

Health and Activity Limitation Survey (HALS)

Ms. Adelle Furrie
Statistics Canada

Ontario Health Status Survey

Mr. David Bogart
Director of Information Resources
and Services Branch
Ontario Ministry of Health

Dr. Lily Eastridge
Epidemiologist
Ontario Ministry of Health

General Availability of Ontario Data Sources

Ms. Jan Kestle
Sr. Policy Advisor
Ministry of Treasury and Economics
Government of Ontario

Ontario Worker Databases

Dr. Peter Pelmear
Health and Safety Support Services Branch
Ontario Ministry of Labour

Dr. Jaan Roos,
Senior Medical Consultant
Chest Clinic, Health and Safety Services Branch
Ontario Ministry of Labour

Dr. Jim Stopps, Chief
Health Studies Services
Ontario Ministry of Labour

Mr. R. Hanna
Systems Specialist
Systems Prototyping Development
and Maintenance
Ontario Ministry of Labour

U.S. Experiences at Linking Administrative and Health Databases

Mr. Peter Sailor
U.S. Internal Revenue Service

Collection of New Occupational Data

Dr. Pierre Band
Director, Division of Epidemiology and Statistics
Cancer Control Agency of British Columbia

Dr. Gerry Hill
Department of Epidemiology and Bioststistics
McGill University
(formerly with Alberta Cancer Board)

Ms. Shirley Fincham
Department of Epidemiology and Preventive Oncology
Alberta Cancer Board

Dr. Jack Siemiatycki
Centre de recherche en epidemiologie
et medecine preventive
Institut Armand-Frappier

Dr. G. Marie Swanson,
Director, Detroit Cancer Registry
Michigan Cancer Foundation

Methodologic Issues

Dr. Neil Pearce (cancer cases as controls)
Department of Community Medicine
Wellington Hospital, School of Medicine
Wellington, New Zealand

Mr. Gilles Montigny (computerized occupation coding)
Labour and Household Survey Analysis Division
Statistics Canada

Mr. Mike Wenzowski (computerized occupation coding)
Project Officer, Labour and Household Survey Analysis Division
Statistics Canada

Mr. Tony Malfara
Standards Division
Statistics Canada

Dr. Karl Sieber (NIOSH job exposure matrix)
U.S. National Institute of Occupational Safety and Health (NIOSH)

Mr. Michael McElroy (U.S. job classification systems)
Supervisory Economist, Bureau of Labour Statistics
U.S. Department of Labour

Mr. John C. Thompson (U.S. job classification systems)
Bureau of Labour Statistics
U.S. Department of Labour

National Crosswalk Conversion Centre (conversion of U.S. job classification systems)
Des Moines, Iowa

Chief, Population Division (U.S. job classification systems)
Census Coding Service
U.S. Bureau of the Census

APPENDIX 3.

Letter from

Mr. Bruce Petrie,
Assistant Chief Statistician,
Social, Institutions, and Labour Statistics Field
Statistics Canada

Statistics Canada Statistique Canada
Ottawa, Canada

File
Chrono
D. Bray

April 15, 1988

Loraine D. Marrett, Ph.D.
Senior Epidemiologist
Division of Epidemiology and Statistics
Ontario Cancer Treatment and
Research Foundation
7 Overlea Boulevard
Toronto, Ontario
M4H 1A8

Dear Dr. Marrett:

I am writing in response to your letter of February 12 and our subsequent meeting and conversations with respect to the feasibility of linking census and Ontario Cancer Registry data to establish an occupational cancer database for the province of Ontario. Following further discussion with my colleagues, I can now advise you that Statistics Canada is prepared to cooperate with the OCTRF in an effort to identify how such a database might be established in a way which would both meet your research objectives and respect Statistics Canada's concerns and requirements with respect to matters of confidentiality and privacy.

As I indicated during our most recent telephone conversation, any approach which might be developed would have to be based on the provision of aggregated data, rather than micro-records. I also noted that several of my colleagues expressed doubts about the likelihood of being able to generate meaningful data in a cost-effective manner, given the nature of the data involved and the manner in which it is stored.

...2

I would suggest, then, that a feasibility study should be carried out to determine whether a mutually satisfactory methodology can be developed. Should the study identify a practical approach, we would then proceed, pursuant to our policy, to seek ministerial approval of the record linkage proposal.

I have asked John Coombs and his staff in Health Division to work with you to pursue this initiative, and indicated that you would be contacting him soon.

Yours sincerely,

D. Bruce Petrie
Assistant Chief Statistician
Social, Institutions and
Labour Statistics Field

x.c. Mr. John Coombs

APPENDIX 4.

Cancer Act of Ontario (Revised Statutes, 1980)

CHAPTER 57

Cancer Act

PART I

THE ONTARIO CANCER TREATMENT AND RESEARCH FOUNDATION

Foundation continued

1. The corporation known as The Ontario Cancer Treatment and Research Foundation, referred to in this Act as the Foundation, is continued. R.S.O. 1970, c. 55, s. 1.

Members

2.—(1) The Foundation shall consist of not fewer than seven members who shall be appointed by the Lieutenant Governor in Council and who shall hold office during pleasure.

Vacancies

(2) The Lieutenant Governor in Council may fill any vacancies that occur from time to time in the membership of the Foundation.

Quorum

(3) Five of the members of the Foundation constitute a quorum for the transaction of business. R.S.O. 1970, c. 55, s. 2.

Chairman, vice-chairman

3.—(1) The Lieutenant Governor in Council may appoint one of the members to be chairman of the Foundation and another of the members to be vice-chairman of the Foundation.

Presiding officer

(2) The chairman shall preside at all meetings of the Foundation at which he is present and in the absence of both the vice-chairman shall preside and in the absence of both the chairman and the vice-chairman the members present shall elect one of themselves to preside. R.S.O. 1970, c. 55, s. 3.

Advisory medical board

4. Subject to the approval of the Lieutenant Governor in Council, the Foundation may appoint an advisory medical board consisting of such persons representative of the medical faculties of the University of Toronto, Queen's University, The University of Western Ontario and the University of Ottawa, and of radiotherapists, surgeons, pathologists, internists, physicists and the medical profession generally as the Foundation considers appropriate. R.S.O. 1970, c. 55, s. 4, *revised*.

Object

5. The object of the Foundation is to establish and conduct a program of research, diagnosis and treatment in cancer, including,

(*a*) the establishment, maintenance and operation of research, diagnostic and treatment centres in general hospitals or elsewhere;

(*b*) the transportation of patients and escorts to its treatment centres or to the hospital of the Institute for diagnosis, treatment or investigation;

(*c*) the establishment, maintenance and operation of hostels in connection with its treatment centres or the hospital of the Institute;

(*d*) the laboratory and clinical investigation of cancer problems;

(*e*) the co-ordination of facilities for treatment;

(*f*) the adequate reporting of cases and the recording and compilation of data;

(*g*) the education of the public in the importance of early recognition and treatment;

(*h*) the providing of facilities for undergraduate and post-graduate study;

(*i*) the training of technical personnel; and

(*j*) the providing and awarding of research fellowships. R.S.O. 1970, c. 55, s. 5.

Agreements

6. Subject to the approval of the Lieutenant Governor in Council, the Foundation may make agreements with universities, medical associations, hospitals and persons for the purpose of carrying out the object of the Foundation. R.S.O. 1970, c. 55, s. 6.

Information to be confidential

7.—(1) Any information or report respecting a case of cancer furnished to the Foundation by any person shall be kept confidential and shall not be used or disclosed by the Foundation to any person for any purpose other than for compiling statistics or carrying out medical or epidemiological research.

Liability

(2) No action or other proceeding for damages lies or shall be instituted against any legally qualified medical practi-

tioner or any licensed dental surgeon or any hospital in respect of the furnishing to the Foundation of any information or report with respect to a case of cancer examined, diagnosed or treated, by such medical practitioner or dental surgeon or at such hospital. 1972, c. 34, s. 1.

Staff

8. The Foundation may employ a director and officers, clerks and servants and may engage the services of experts and other persons and may pay such director, officers, clerks, servants, experts or other persons such remuneration as it considers proper out of its funds. R.S.O. 1970, c. 55, s. 7.

By-laws

9. Subject to the approval of the Lieutenant Governor in Council, the Foundation may make such by-laws, rules or regulations as are considered expedient for the administration of its affairs. R.S.O. 1970, c. 55, s. 8.

Funds

10. The funds of the Foundation consist of moneys received by it from any source including moneys appropriated for its use by the Parliament of Canada or the Legislature of Ontario, and the Foundation may disburse, expend or otherwise deal with any of its funds in such manner not contrary to law as it considers proper. R.S.O. 1970, c. 55, s. 9.

Expenses

11. The members of the Foundation and its medical advisory board shall be paid such amounts for travelling and other expenses as the Foundation, subject to the approval of the Lieutenant Governor in Council, may determine from time to time. R.S.O. 1970, c. 55, s. 10.

Audit

12. The accounts of the Foundation shall be audited annually by the Provincial Auditor or by such qualified auditor as the Lieutenant Governor in Council designates, in which event the costs of the audit shall be paid out of the funds of the Foundation. R.S.O. 1970, c. 55, s. 11.

Annual report

13.—(1) The Foundation shall after the close of each fiscal year make a report upon its affairs during the preceding year to the Minister of Health and every such report shall contain a financial statement, certified by the auditor, showing all moneys received and disbursed by the Foundation during the preceding year. R.S.O. 1970, c. 55, s. 12 (1).

Idem

(2) The Minister of Health shall submit the report to the Lieutenant Governor in Council and shall then lay the report before the Assembly if it is in session or, if not, at the next ensuing session. R.S.O. 1970, c. 55, s. 12 (2); 1972, c. 1, s. 78 (1).

Power to expropriate land and erect buildings

14.—(1) Subject to the approval of the Lieutenant Governor in Council, the Foundation may acquire by purchase or lease, or may enter upon, take and use without the consent of the owner thereof, any land and buildings that are considered suitable for the purposes of the Foundation and may erect buildings, acquire and install machinery and equipment and purchase all such instruments, materials and appliances and other matters and things that are considered necessary.

Application of R.S.O. 1970, c. 148

(2) Whenever the Foundation exercises the power to enter upon, take or use lands without the consent of the owner thereof, the *Expropriations Act* applies. R.S.O. 1970, c. 55, s. 13.

Right to acquire patents, etc.

15. Subject to the approval of the Lieutenant Governor in Council, the Foundation may apply for, or acquire by purchase, assignment or otherwise, rights in any patent relating to any remedy for the prevention or cure of cancer and may sell and dispose thereof or of any interest therein, and grant or assign any rights that have been acquired by the Foundation thereunder. R.S.O. 1970, c. 55, s. 14.

Property not liable to assessment

16. The real and personal property, business and income of the Foundation is not subject to taxation for municipal or provincial purposes. R.S.O. 1970, c. 55, s. 15.

PART II

THE ONTARIO CANCER INSTITUTE

Institute continued

17. The corporation known as The Ontario Cancer Institute, referred to in this Act as the Institute, is continued. R.S.O. 1970, c. 55, s. 16.

Members

18.—(1) The Institute shall consist of fifteen persons appointed by the Lieutenant Governor in Council, namely,

(*a*) five persons representing the Foundation, one of whom shall be the chairman of the Foundation;

(*b*) three persons representing The Governing Council of the University of Toronto;

(*c*) one person representing the Board of Trustees of the Toronto General Hospital;

(*d*) one person representing the Board of Trustees of The Hospital for Sick Children;

(*e*) one person representing the governing body of St. Michael's Hospital;

(*f*) one person representing the Board of Governors of The Toronto Western Hospital;

(*g*) one person representing the Board of Governors of the Women's College Hospital;

(*h*) one person representing the Board of Governors of the Toronto Wellesley Hospital;

(*i*) one person representing the Board of Governors of New Mount Sinai Hospital,

who shall hold office during pleasure.

Vacancies

(2) The Lieutenant Governor in Council may fill any vacancies that occur from time to time in the membership of the Institute in accordance with the method of representation prescribed in this section.

Quorum

(3) Five of the members of the Institute constitute a quorum for the transaction of business. R.S.O. 1970, c. 55, s. 17, *revised*.

Chairman

19. The Lieutenant Governor in Council may appoint one of the representatives of the Foundation as chairman of the Institute. R.S.O. 1970, c. 55, s. 18.

Advisory medical board

20. Subject to the approval of the Lieutenant Governor in Council, the Institute may appoint an advisory medical board consisting of legally qualified medical practitioners, scientists and other persons R.S.O. 1970, c. 55, s. 19.

Object

21. The object of the Institute is to maintain, manage and operate a provincial hospital with facilities for cancer research, diagnosis and treatment. R.S.O. 1970, c. 55, s. 20.

Agreements

22. Subject to the approval of the Lieutenant Governor in Council, the Institute may make agreements with the Foundation or with any university, medical association, hospital or person for the purpose of carrying out the object of the Institute. R.S.O. 1970, c. 55, s. 21.

Staff

23. The Institute may employ a director and such staff as may from time to time be required for the purposes of the hospital and may pay such director and staff such remuneration as it considers proper out of its funds. R.S.O. 1970, c. 55, s. 22.

By-laws

24. Subject to the approval of the Lieutenant Governor in Council, the Institute may make such by-laws, rules or

regulations as are considered expedient for the administration of its affairs. R.S.O. 1970, c. 55, s. 23.

Funds

25.—(1) The funds of the Institute consist of moneys received by it from any source, including the Foundation, and the Institute may disburse, expend or otherwise deal with any of its funds in such manner not contrary to law as it considers proper.

Estimates

(2) The Institute shall annually prepare and submit to the Foundation the estimates of the moneys required for its purposes during the ensuing fiscal year. R.S.O. 1970, c. 55, s. 24.

Expenses

26. The members of the Institute and its medical advisory board shall be paid such amounts for travelling and other expenses as the Institute, subject to the approval of the Lieutenant Governor in Council, determines from time to time. R.S.O. 1970, c. 55, s. 25.

Audit

27. The accounts of the Institute shall be audited annually by the Provincial Auditor or by such qualified auditor as the Lieutenant Governor in Council designates, in which event the costs of the audit shall be paid out of the funds of the Institute. R.S.O. 1970, c. 55, s. 26.

Annual report

28.—(1) The Institute shall after the close of each fiscal year make a report upon its affairs during the preceding year to the Minister of Health and every such report shall contain a financial statement, certified by the auditor, showing all moneys received and disbursed by the Institute during the preceding year. R.S.O. 1970, c. 55, s. 27 (1).

Idem

(2) The Minister of Health shall submit the report to the Lieutenant Governor in Council and shall then lay the report before the Assembly if it is in session or, if not, at the next ensuing session. R.S.O. 1970, c. 55, s. 27 (2); 1972, c. 1, s. 78 (2).

Property not liable to assessment

29. The real and personal property, business and income of the Institute are not subject to taxation for municipal or provincial purposes. R.S.O. 1970, c. 55, s. 28.

APPENDIX 5.

Description of the Ontario Cancer Registry

from

"The Ontario Cancer Registry:
Twenty Years of Cancer Incidence, 1964-1983"

L. D. Marrett

Twenty Years of Cancer Incidence 1964-1983 The Ontario Cancer Registry

E.A. CLARKE
L.D. MARRETT
N. KREIGER

November 1987

THE ONTARIO CANCER TREATMENT AND RESEARCH FOUNDATION

Cancer Registration in Ontario

Introduction

This monograph is the first of a proposed series on cancer incidence, survival and mortality in Ontario. The occurrence of cancer is presented by year, site, sex and age for twenty years in Ontario. This information has not been published in a consolidated form elsewhere and supercedes data previously published for selected time periods.

The monograph is intended as a reference for researchers, health care workers and others. A description of the unique method of cancer registration employed in Ontario precedes the presentation of the data. An understanding of the techniques employed is necessary for interpretation of the data and, where necessary, the reader is cautioned as to possible limitations and artifacts.

Background

Ontario is the most populous province in Canada, with 9.1 million people in 1986[1] and an area of over one million square kilometers between Quebec to the east, Manitoba to the west, Hudson's Bay to the north and the United States to the south. Located between latitudes 41° - 57° north and longitudes 74° - 95° west, the altitude varies from sea level to 690 metres. Although 82% of the population now inhabit urban areas, mostly in the southern part of the province, previously there have been sizeable farming and mining communities. Manufacture of both durable and nondurable goods, trade and communications are the major industries. Eighty percent of Ontario residents were born in Canada but nevertheless represent a wide variety of ethnic groups; 60% are of British origin, 8% of French origin, 6% of Italian origin and 5% of German origin. In addition, there are about 84,000 native people in the province. The rate of emigration from Ontario is low. Of all persons residing in Ontario in 1976, only 4.0% had left the province by 1981. The rate of emigration for those more than 45 years of age, who are at greater risk of cancer, is even lower (1.4%)[2].

The Ontario Cancer Registry (OCR) is a population-based registry covering the entire province of Ontario. It is operated by the Ontario Cancer Treatment and Research Foundation, which was incorporated in 1943 by an Act of the Legislature of the Province of Ontario (The Cancer Act) "to establish a programme of cancer diagnosis, treatment and research" in the province. This act followed a recommendation by a provincial commission (the Cody Commission), that radiotherapy, then the most effective method of cancer treatment other than surgery, be centralized. Regional cancer centres (RCC) were therefore established to provide radiotherapy to outpatients and also included research activities. RCC are now located in Hamilton, Kingston, London, Ottawa, Sudbury, Thunder Bay, Windsor and Toronto. In addition, the Ontario Cancer Institute, incorporating the Princess Margaret Hospital (PMH), was established by statute in 1952 and in 1958 opened in Toronto with two research divisions. The PMH is the only hospital in Ontario dedicated exclusively to the care of cancer patients. Together, the RCC and the PMH provide all the radiation therapy for cancer patients in the province, as well as chemotherapy and consultative services for approximately 50% of cancer patients in Ontario.

The Ontario Cancer Treatment and Research Foundation, including the OCR, is supported primarily by the Ontario Ministry of Health (MOH). Patient care is publicly financed; in Ontario about 95% of Ontario residents are covered by a comprehensive government health insurance plan. While some residents of Ontario seek medical care outside the province, the proportion of claims for in-patient care originating from outside Ontario is less than 1%. The majority of such claims are made by residents of Ontario who live close to its borders.

The Cancer Act of 1943 included provision for "the adequate reporting of cancer cases and the recording and compilation of data". Cancer is not a legally reportable disease in Ontario, but amendments to the Cancer Act since 1943 have provided legal protection for organizations or individuals in the health care system who report information on cases of cancer to the Ontario Cancer Treatment and Research Foundation. These amendments enable information in the OCR to be used for epidemiological and medical research. In addition, each hospital in the province is required to forward diagnostic information on every discharged patient to the MOH. The MOH uses this information for administrative purposes and provides the OCR with copies of data on cancer patients; thus, a degree of compulsory reporting is in effect for hospitalized patients.

Establishment of the Ontario Cancer Registry

In 1936 Dr. A.H. Sellers, at that time with the Department of Health for Ontario - forerunner of Ministry of Health of Ontario - recognized that accurate data on the incidence and prevalence of cancer in Ontario were desirable and necessary. As in the United States, periodic surveys of cancer incidence were then undertaken. The first such survey took place in 1939. Physicians and hospitals in one county of Ontario registered each case of cancer seen in 1939, by completing a form giving residence, age, site and histology[3]. This survey was repeated in the same region in 1953[4].

In 1963 the Ontario Cancer Treatment and Research Foundation considered various methods of obtaining information necessary for the establishment and operation of a population-based cancer registry for Ontario. It was decided to use existing and available information contained in reports routinely prepared for other purposes, without requiring legislation that cancer be made a reportable disease.

Initially, a manual system was established in which information from the following sources was recorded:

- hospital in-patient separation forms in which cancer was given as a diagnosis.

- abstracts of death registrations in which cancer was the underlying cause of death, as well as those in which cancer was mentioned.
- abstract cards used by the RCC and the PMH to record information on their cancer patients.
- reports from cancer registries maintained by the RCC.
- reports from hospital-based cancer registries.
- reports on biopsies performed on outpatients with cancer.
- drug prescriptions for needy cancer patients.
- out-patient radiotherapy reports.

After abstraction, coding and transcription of the information from the above sources to cards at the OCR, a manual linkage based on surname and given name was performed to bring together all records for an individual. The primary site of cancer was determined by a physician reviewing all records for each individual. Incidence data for the entire province were published for 1964 in 1968[5], for 1965 in 1970[6] and for 1966 in 1973[7].

The lack of timely incidence data indicated the necessity to streamline the process by implementing computer techniques. It was anticipated that this would permit a small cohesive group to handle a large number of records in a consistent manner and thus generate incidence data that were timely and of high quality.

Sources of Data

There have been major changes over time in the sources and quality of the data used for input to the Registry. The numbers of reports from the various sources used by the OCR are shown in Table A. Although many different sources were employed in the early years, the OCR has relied on only four major sources since 1972. They are:

- hospital separations with cancer as a diagnosis.
- pathology reports with a mention of cancer.
- death certificates in which cancer was the underlying cause of death.
- reports on patients referred to the RCC and PMH.

Hospital Discharges

Hospital inpatient separation data with mention of cancer are forwarded to the OCR by the MOH. These were submitted as documents until 1975, after which time the data were provided on magnetic tape. In 1978 the MOH instituted a requirement that each hospital submit an abstract for each discharge to an independent organization, the Hospital Medical Records Institute (HMRI). The HMRI abstract form provides for the recording of sixteen possible discharge diagnoses (as opposed to the single diagnosis previously permitted on hospital separation forms) but these abstracts do not contain surnames or given names. After processing (which includes some editing), HMRI forwards the resulting file to the MOH where name and Ontario Health Insurance Plan (OHIP) number are added. A subset of this integrated file is created consisting of records in which cancer is one of the discharge diagnoses and this file is forwarded annually to the OCR. The number of reports received in 1964 included patients whose admission to hospital occurred in 1963 or earlier. From 1965 onwards the number of cancer records submitted by hospitals has increased each year, with the exception of 1975 when the MOH changed from a manual to a computer system. The large increase in the number of records in 1978 is attributed to the capture of any record from the HMRI file in which any one of the sixteen diagnoses was cancer. In 1983, the increase is attributed to the inclusion of records in which a history of cancer is given.

Pathology Reports

In 1973 pathology laboratories across the province were asked to submit copies of reports in which cancer was mentioned. By 1980 all were complying. Paper records are provided to the OCR by participating laboratories and are coded by OCR staff.

Deaths

The OCR has data in machine-readable form on all deaths in Ontario residents since 1964. For the years 1964-1980, these data were received from Statistics Canada, by special arrangement with the Office of the Registrar General of Ontario. Since 1981, the Office of the Registrar General of Ontario has annually provided a computer tape directly to the OCR. Underlying cause of death is coded by trained nosologists in the Office of the Registrar General. Deaths with cancer considered to be the underlying cause have been included in the OCR since 1964. In addition, abstracts of death certificates with mention of cancer, provided by the Office of the Registrar General, have been included for the years 1964-1971.

Treatment Centres and Registries

Initially, abstract cards recording minimal information on their cancer patients were completed at each RCC and the PMH. Those from the RCC were forwarded to the OCR for further data abstraction and coding. Between 1964 and 1981, these cards were gradually discontinued at the PMH and the RCC and appropriate data were subsequently forwarded to the OCR in machine-readable form. Abstract cards were also created for tumour registries maintained at the RCC for cases diagnosed in their regions but not referred to the centres. These cards were forwarded to the OCR for abstracting and coding until the registries were discontinued by the RCC.

Other Sources

Pathology reports from hospital laboratories across the province duplicated the information contained in the biopsy reports on outpatients and were more comprehensive. The number of hospital-based registries in the province diminished due to financial constraints and use of these two sources by the OCR was discontinued in 1972. When incidence data from 1969 to 1971 were first generated and reviewed, it was realized that no cases were identified by radiotherapy out-patient reports per se and only 0.3% solely by the Foundation's Drug Service Programme. The identifying information on these sources was poor and it was decided to discontinue their use in 1972.

TABLE A

Reports Used in Registering Cases by Year and Source

Year	Hospital Discharge	Pathology	Death	RCC/PMH	RCC Registry	Hospital Registry	Biopsy Service	Drug Service	Outpatient Radiotherapy	Total
1964	23226	10	8831	6439	481	333	429	162	3	39914
1965	19033	11	9154	6534	984	864	366	206	2	37154
1966	21710	196	9380	6643	1143	836	541	191	18	40658
1967	25918	29	9849	7002	1285	3961	114	334	22	48514
1968	34974	11	11459	7127	1164	3901	13	605	6082	65336
1969	43411	158	11802	7736	2133	3565	1215	506	4456	74982
1970	40916	150	12191	8011	2291	31	449	2373	5858	72270
1971	42768	454	12203	8924	2113	24 [1]	259 [1]	481 [1]	6552 [1]	73778
1972	49699	908	11577	9610	2181	-	-	-	-	73975
1973	56153	11941	11794	9862	1703	-	-	-	-	91453
1974	60005	12519	12088	10720	2289	-	-	-	-	97621
1975	57350	13345	12224	11194	1974	-	-	-	-	96087
1976	61358	15032	12533	11997	1090 [2]	-	-	-	-	102010
1977	63430	15885	13100	12169	-	-	-	-	-	104584
1978	76870	20717	13428	14402	-	-	-	-	-	125417
1979	80257	26495	13961	13851	-	-	-	-	-	134564
1980	84056	28192	14217	14632	-	-	-	-	-	141097
1981	86792	34147	14854	15303	-	-	-	-	-	151096
1982	87098	35955	15306	16139	-	-	-	-	-	154498
1983	99991	43740	15869	17597	-	-	-	-	-	177197
Total	1115015	259895	245820	215892	20831	13515	3386	4858	22993	1902205

[1] Use of these records was discontinued after 1971.
[2] Use of these records was discontinued after 1976.

Coding, Data Entry and Preprocessing of Data

All cancer records submitted to the OCR in the early years (1964-1975), except death records in which cancer was reported as the underlying cause, were coded and entered into the computer centrally by the OCR. Between 1975 and 1977, hospital discharge information was coded at the MOH and from 1978 it has been coded in the medical records departments of hospitals in Ontario. These data have been sent to the OCR on magnetic tape since 1975. Given the fact that a passive system of cancer registration is employed, it is not possible, for the most part, to institute formal methods of quality control in regard to coding.

Pathology reports have always been coded and the data entered by clerks at the OCR. These are subjected to routine assessment of quality, as were other records previously coded at the OCR. Difficult reports are circulated among coding staff and are discussed at regular meetings with the medical staff.

Data from the RCC and PMH have been collected uniformly since their establishment. With computerization of records at these centres, coding has devolved to their medical record staffs. The Managers of Health Records at each RCC and the PMH meet twice a year to discuss coding and other quality control issues. The RCC also send copies of pathology reports and a clinical description of the cancer to the OCR. These reports are recoded, and any discrepancies are corrected after discussion between the RCC and the OCR.

Routine quality control of the data entry phase is carried out on all records of the OCR. Samples of reports entered on-line are verified by routine recoding and key entry. The data entry system requires that certain variables (e.g., surname of patient, date of diagnosis, site of disease) always be entered. As data are entered, they are edited for validity, consistency and plausibility. Data received on magnetic

tapes are also subjected to the same edits; however, these are carried out by batch programmes. Validity edits reject data which are inherently incorrect (e.g., the 13th month, the 32nd day). Consistency edits compare two or more data fields and report contradictions (e.g., a male patient with ovarian cancer, a treatment date preceding date of birth). Edits for plausibility report unlikely but possible situations which are potential errors (e.g., a 110 year old patient, a five year old male with prostatic cancer). These plausibility edits are checked manually and corrected if necessary. Coded data (e.g., residence, hospital, birthplace) are compared with tables constructed by the OCR specifically for validation purposes. Finally, numerical data are validated with check digits.

Site of cancer in abstracts of death certificates for the years 1964-1968 was originally coded to the Seventh Revision of the International Classification of Diseases (ICD)[8], and for 1969-1978 to the Eighth Revision[9]. From 1964 until 1979, site of cancer on all other records in the OCR was coded to the Eighth Revision. Site on all records since 1979 has been coded to the Ninth Revision of the ICD (ICD9)[10]. In addition to the computer edits described above, all ICD codes are converted during processing to the ICD9. Prior to 1979, morphology was coded to the Manual of Tumor Nomenclature and Coding (MOTNAC)[11] and, since 1979, to the International Classification of Diseases for Oncology (ICDO-M)[12]. MOTNAC codes are also converted to ICDO-M codes. All cases in the OCR are reported in ICD9 and ICDO-M.

Creation of Cancer Incidence Data

The methods of generating the cancer incidence data reported in this monograph are essentially the same over the twenty year time period and result from a collaboration between two departments of the Ontario Cancer Treatment and Research Foundation, the Division of Epidemiology and Statistics and Information Systems.

Once the source files have been preprocessed, cancer incidence data are created in two major phases. First, sequential computer linkage of the source files is carried out to bring all records pertaining to an individual together. Then, a set of computerized rules known as the "Case Resolution System" is applied to the linked records, to allocate the appropriate site of disease, histology, date and method of diagnosis, residence, and other information for each case of cancer.

Linkage of Data Sources

As described earlier, all records from 1964-1966 were linked manually. These linked records, together with the source data for 1967-1971, were then keypunched and stored in machine-readable form. The manual linkage of source records was replaced by computer programmes written by OCR staff to identify records having a high probability of belonging to the same individual, using record linkage principles developed by Newcombe, Kennedy and Smith[13,14,15]. The entire output from this record linkage of 1964-1971 data was manually edited to match records that were not linked by the computer and to separate those records that were linked incorrectly. A master record was created for each case and then inspected to determine whether the site, histology and date of diagnosis had been correctly assigned; these were then modified, if indicated. This simple linkage system and detailed manual editing were very labour intensive, again resulting in a delay in the production of incidence data. Cancer incidence data for 1969 to 1971 were not published until 1978[16].

Fortunately, by 1978 more efficient computers and sophisticated computerized record linkage systems were available which facilitated the task of linking large numbers of records. In order to link together the large volume of data, OCR staff have developed a sophisticated computer record linkage system based on the Generalized Iterative Record Linkage System (GIRLS)[17] designed by Statistics Canada in conjunction with the Epidemiology Unit of the National Cancer Institute of Canada.

Since Ontario does not have a unique number in the health or political system which identifies an individual throughout life, linkage is based on a number of identifying variables including name, date of birth, OHIP number, hospital where diagnosed and hospital chart number. It should be noted that an OHIP number is allocated to a family, and does not distinguish between individual members of that family. When a child reaches the age of 18 (or 21, if attending university), he or she is assigned a new OHIP number. Change of employment, or divorce, may also result in the allocation of new OHIP numbers to individuals.

The present computer linkage is completed in several stages. First, a NYSIIS (New York State Intelligence Identification System) code is created which is a phonetic version of the surname. Only records which have the same NYSIIS code are compared for possible linkage; therefore, records with names having similar spellings but different NYSIIS codes do not have an opportunity to link. Records with the same NYSIIS code constitute a "pocket" within which records are compared. A numeric score or "weight" is assigned to each variable when two records are compared. The greater the sum of the weights of the variables compared, the greater the probability that two records linked by the system belong to the same individual. The word "iterative" in the acronym GIRLS indicates that this process of allocating weights is repeated more than once. The system uses previous observations to assign more precise weights.

Each link (i.e., each pair of records brought together) is classified into one of three categories: definite, possible or rejected, based on the magnitude of the total weight. The distribution of the total weights in the linked file is usually bimodal, clustering around a high weight (definite, i.e., likely to be true links) and a low weight (rejected, i.e., unlikely to be true links). The middle range of weights contains possible links, i.e., those in which it is uncertain that paired records relate to the same individual. Linked records in this range (the grey area) are reviewed by health record staff of the OCR who have access to additional data that were not used in the linkage. An example would be information contained in the complete pathology report which might confirm the suspicion by the staff that an earlier biopsy had

been performed. Decisions are made to accept or to reject each link in the grey area and the result is then entered into the linked files. This manual resolution reduces the number of false links accepted and missed true links, but both still occur. The size of the grey area varies according to the files being linked; between 2 and 12% of potential links are manually resolved.

Linkages of source files are performed in sequence (Figure 1). Each year, hospital reports are linked internally to bring together multiple admissions for the same patient. Pathology reports are then linked to these aggregated hospital records, since most pathology reports will be related to a hospital stay. This combined hospital-pathology file is subsequently linked with previous years' incidence to identify incident (as opposed to prevalent) cases, producing provisional incidence data. Every second or third year, deaths due to cancer and records from the RCC and the PMH are linked to these provisional data. These final linkages add few new cases of cancer, although RCC and PMH records improve data quality, particularly the specificity of site and histology. This sequence of linkages has been applied to data from 1972 to the present. All RCC, PMH and death records from 1964 to 1983 have been included in the data presented in this monograph.

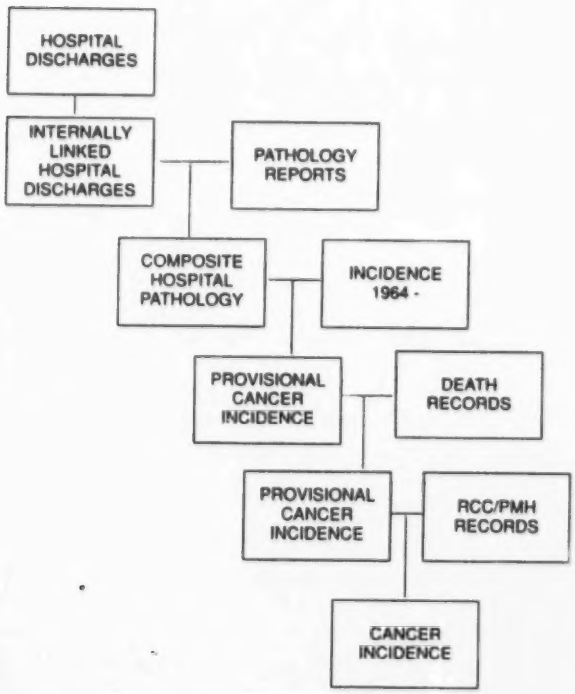

Fig. 1 Sequence of linkages of source files for generation of cancer incidence data.

An internal linkage has been performed on the entire 1964-1983 file using pockets other than those created by NYSIIS codes. This linkage allowed groups of records with different NYSIIS codes to be compared so that records which may have pertained to the same patient had the opportunity to link. There were three distinct phases to this linkage. Within pockets created in each phase, comparisons of all possible pairs of records were carried out and weights were assigned, as in other linkages. In the first phase, pockets were assigned using OHIP number and sex. In the second phase, pockets were formed using birth year and the first three letters of the given name. The third phase utilized the first three characters of the surname. Records linking at a high weight in one phase were not included in subsequent phases. The grey area resulting from this three-phase linkage was resolved as in other routine linkages. As a result of this linkage, 2.5% of patients were identified as being recorded twice in the OCR and records for these patients were amalgamated. This reduced the effect on the OCR due to errors in spelling or in transcription of surnames.

Nearly all cancer patients have multiple source records due to this method of cancer registration. A set of computer programmes has been developed by OCR staff to create a composite identification record containing the "best" identifying information as determined from all source records on a patient (e.g., surname, given names, birthdate, sex). This is then carried forward into the next linkage. These programmes also find conflicts between individual source records which may be the result of false links which had not been identified earlier. These conflicts are reported and reviewed by OCR health record staff, who make corrections as indicated.

Results of the early manual linkages have been set aside, original 1964-66 source records have been computerized and the same linkage techniques and rules have been applied over the twenty years of data reported in this monograph. As a result, data reported previously for earlier years[5,6,7] are not the same as those reported here.

Allocation of Site, Histology and Other Information

Groups of linked source records for individual patients are processed by a second major system, "Case Resolution". This consists of a series of computer modules developed by OCR staff, and applies medical logic to the source records for a patient to determine the appropriate site of disease, histology and date of diagnosis, since these may vary between source records.

In the Case Resolution system, cancer sites on all source records belonging to one individual are examined to determine the most specific site in a rubric or group of rubrics of the ICD. Only the most specific site codes are retained for further processing. Thus, if one source record indicated "malignant neoplasm of digestive tract" (ICD9 159.9), another indicated "malignant neoplasm of stomach not otherwise specified" (ICD9 151.9), another "malignant neoplasm of pylorus" (ICD9 151.1) and another "malignant neoplasm of plyoric antrum" (ICD9 151.2), only codes ICD9

151.1 and ICD9 151.2 would be retained because they are the most specific.

If at this stage only one site code remains, it is deemed to represent the primary site. If more than one remains but the only difference occurs in the 4th digit of the ICD9 (e.g., 151.1 and 151.2), then the site is selected from the most reliable source. For this purpose, RCC and PMH records are considered to be the most reliable source, followed by pathology records, then hospital discharge records and, finally, death certificates.

If more than one 3-digit site code remains, histology codes on each record for a patient are compared. Histology codes considered to be the same are organized into groups, according to a modification of the classification prepared by Berg[18], and are presented in Table B. Records with a blank histology field, or in which the histology is either "neoplasms not otherwise specified" (ICDO-M 800) or "no microscopic confirmation of tumour" (ICDO-M 999), are included in all histology groups. In addition, records in which the histology given is "carcinoma not otherwise specified" (ICDO-M 801) and "carcinoma undifferentiated type not otherwise specified" (ICDO-M 802) are included with all histology groups except sarcoma, lymphoma and leukaemia. The ICDO-M codes not given in the table are considered to each have a different histology from any other, for example, "mucoepidermoid neoplasms" (ICDO-M code 843).

The OCR considers a second site of cancer in the same individual to be metastatic unless clearly shown to be

TABLE B

Groupings of histology codes considered the "same" for allocation of site in the OCR

Name ICDO-M	Code ICDO-M
Squamous cell carcinomas	807-808
Transitional cell carcinomas	812-813
Adenocarcinomas	814,816,818-823, 825-838,857
Adnexal carcinomas	839-842
Cystic, mucinous and serous carcinomas	844-848
Ductal carcinomas	850-854
Specialized gonadal carcinomas	859,860,862-867
Paragangliomas and glomus tumours	868-871
Melanomas	872-874,876-878
Sarcomas and other soft tissue tumours	880,881,883-886, 889-892,899
Teratomatous carcinomas	908,909
Blood vessel and lymphatic vessel carcinomas	912-915,917
Osteosarcomas, chondrosarcomas and odontogenic tumours	918,919,922-927,929-933
Other tumours (pinealoma, chordoma and granular cell myoblastoma)	936,937
Gliomas	938-948
Neuroepitheliomatous tumours	949,950
Nerve sheath tumours	954,956
Lymphomas and Hodgkin's disease	959-966,969-972,975
Leukaemias	980-994

otherwise. Thus the rules for reporting second primary cancers are conservative. For two different primary sites to be reported in the same individual, the sites must be different at the 3-digit ICD9 level and the histologies of the two sites must be in different groups, as given in Table B. The only exception to this rule is breast cancer. Other sites rarely metastasize to the breast; if breast cancer is given as the site on a source record, then it will always be reported as a primary site, even if the histology is in the same group as that of other primary sites in the linked records. The case resolved from the other primary sites is also reported.

If different 3-digit site records have histologies in the same group according to Table B, one or more sites are considered to be metastatic and the site reported by the most reliable source, as defined earlier, will be allocated as the primary site. If the sources are equally reliable, a broad code which encompasses all the more specific site codes from the most reliable sources is selected as representing the primary site. For example, adenocarcinomas of the transverse colon (ICD9 153.1) and of the rectum (ICD9 154.1), reported by equally reliable sources, would be allocated to "large intestine, not otherwise specified" (ICD9 153.9).

Although the OCR does not report non-melanotic skin cancer (ICD9 173), records of this site are used, if appropriate, to override the site and histology given by less reliable sources. Thus, a pathology record with a diagnosis of cancer of skin of lip (ICD9 173.0) will result in a non-reported case, even although the hospital record indicated a diagnosis of cancer of the lip (ICD9 140.9) as the site of cancer.

The primary site is resolved to "malignant neoplasm without specification" (ICD9 199) when more specific allocation cannot be achieved. This can happen in three ways: first, if the only site recorded is 199; second, if there are two possible primaries in different organ groups with the same histology and equally reliable sources; and third, if more than one secondary site (ICD9 196-198) is specified in the absence of a primary site.

If more than one primary site is identified, each source record is examined and associated with the most appropriate of these diagnoses, thereby aggregating all data for each site. Finally, a composite case record for each case diagnosed is created comprising the best set of diagnostic information, as determined by the computer, using all source records. Checks are made to ensure consistency of composite information, (e.g., that site and histology are not in conflict, that date of death does not precede date last known alive).

The Case Resolution system is being continually improved. Cases which cannot be resolved by the rules, or which include inconsistencies, are reviewed by OCR staff and may result in subsequent modification of the computer rules. Rule changes can be encoded into the system and the entire registry file reprocessed according to the new rules.

Variables in the Ontario Cancer Registry

Three kinds of records exist in the registry: source records, composite identification records and composite case records. Source records are obtained from the external sources previously described, and represent cancer-related events. Composite identification records are created by record linkage, and represent the best identifying and demographic information on each cancer patient. Finally, composite case records are created by the Case Resolution system, and each composite case record describes an individual primary case of cancer. Most analyses use the variables in the composite records as listed in Table C.

Advantages and Limitations of the Ontario Cancer Registry

The advantages of the unique system of registration in Ontario are several. The multiple sources of data combine to provide incidence data of good quality and completeness. A recent study of completeness of cancer registration in 1982, using capture-recapture methodology, estimated completeness for all sites combined as more than 95%, with a low of 91% for cutaneous malignant melanoma to a high of over 98% for deep-seated digestive organs[19]. When data from several sources are present for a given diagnosis, the OCR system takes advantage of the known strengths of each particular source to select the best information for each variable.

The multiple sources generate a "patient profile", in that data from all hospital discharges related to the cancer diagnosis are stored in the OCR, so that length of stay and other data which are valuable to health planners are available. The patient profile also readily permits identification of multiple cancers in the same patient.

TABLE C

Variables Retained in the Composite OCR Record

Composite Identification Record

- registry identification number.
- names (including alternates).
- sex, date and place of birth.
- hospital and residence codes.
- last known date.
- vital status as of last known date.
- cause of death from the death certificate, if deceased.

Composite Case Record

- cancer site and histology codes.
- date of diagnosis.
- method of confirmation.
- residence at time of diagnosis.
- earliest known treatment date.
- hospital and RCC/PMH chart numbers.
- hospital of diagnosis.

This unique method of registration is relatively inexpensive, an important feature in a jurisdiction the size of Ontario.
Finally, the use of computerized linkage and Case Resolution systems ensures that records are processed in a consistent fashion. If the rules for allocating primary site are enhanced, the quality of the incidence data for the entire period of the OCR can be improved, as the complete registry data base (more than 6 million records on more than 500,000 cases) can be processed by the improved system. In addition, the impact of different rules for multiple primaries on incidence rates could be assessed by processing the entire registry file through two separate case resolution programs and comparing the results.
There are, however, some limitations of the system as well. These are primarily related to reliance on, and therefore limited control over, the type, quality and flow of input data. For some data sources, coding of site, histology and residence is decentralized to hospital medical record departments (hospital separation reports) and the Office of the Registrar General (death certificates). Thus, the OCR has no control over the quality of data from these sources, both in coding and in the number of records received. Nevertheless, approximately 50% of cancer cases in the province are eventually referred to the RCC or the PMH, where data quality and uniformity of coding are ensured.
Changes in any of the input data sources can affect registration, so that time trends in cancer incidence may be subject to artifacts related to changes in or problems with sources. For example, in 1978, changes in the administrative arrangements concerning provision of hospital discharge data to the MOH resulted in increased numbers of hospital discharge reports being received by the OCR, thereby producing a sudden increase in incidence rates. Such effects are likely to become less frequent as the OCR matures. Another problem with dependence on outside data sources is the resultant delay in generation of incidence data. HMRI processes all hospital discharge data for a fiscal year at the end of the year; only afterwards are data passed to the OCR. Since hospital discharges comprise the major registry source, linkages cannot begin until these data are received. Nine months of processing are required at the OCR to add one year of data. Therefore incidence data for a particular year are not available until 18-24 months after the close of that year.
Throughout the 1970's the annual number of hospital pathology reports voluntarily submitted to the OCR increased dramatically. Thus, during these years, the OCR would be expected to include increasing numbers of patients reported solely by pathology departments and never admitted to hospital or referred to the PMH or the RCC. For some sites, particularly those in which admission to hospital as an in-patient is not common, artifactual increases may thus be evident throughout this period.
Further problems, occurring from the use of data generated for purposes other than cancer registration, are lack of complete demographic/geographic/tumour-specific information. For example, data are not available on clinical stage at diagnosis except for a proportion of patients seen at the RCC or PMH. Municipality of residence is not available historically for many patients or, currently, for those for whom pathology reports only are received. Age and exact date of birth, an important linkage variable, are sometimes missing, particularly from pathology reports. This contributes to the relatively high proportion (0.06%) of cases reported with unknown age.
In addition, it is likely that there is some over-reporting in the OCR. Although a large proportion of duplicate records (where records for the same individual have not been brought together by the linkage system) were eliminated by the internal linkage process using pockets other than NYSIIS, it is estimated that 0.2% duplicates remain. This is an insoluble problem in a province where individuals do not have unique identifiers. However, the magnitude of the problem of over-reporting, due to failure to correctly link all records, is much less than it would have been if a completely manual linkage were performed. Comparison of data using the present system, published herein, with results of the original 1965-1966 manually generated data[6,7] demonstrates an 11% reduction in the number of cases. Similarly, there is a 5% reduction in the number of cases for 1969-1971 when incidence was generated using simplified computer linkage techniques[16]. When hospital records were coded at the OCR, control of quality and consistency of the coding could be conducted. Currently, as these source records are submitted to the OCR in machine-readable form, control is not possible. This has led to some over-reporting in that hospitals do, on occasion, report cancer cases, as cancer was the admitting diagnosis, whereas the tumours are found upon further investigation in hospital to be non-malignant. The OCR is constantly reviewing sites and age groups where this may occur and taking corrective measures.

Conclusion

Although data for 1964-1982 were generated in large segments, (those for 1972-1976 becoming available in September 1983 and those for 1977-1982 in late 1984), incidence rates since that time have been generated annually. Currently, incidence data are not available for a particular year until 18-24 months after the close of that year. It is expected that this interval will shorten with increasing computerization of hospital discharge and pathology reports at source.
Additional sources, such as cytopathology and haematology reports, may be added to the routine data sources; private pathology laboratories began submitting reports in 1984. This should improve both completeness and data quality. In addition, the OCR is continually investigating new sources of data which could be used to improve the quality of the OCR. For example, a linkage with a government data source is in progress now to improve residence information.
The linkage and Case Resolution systems are constantly being improved and streamlined, and the quality and timeliness of OCR incidence data will improve as they do.
Although the OCR contains twenty years of data, it is a new registry. Only through critical use of the data and comments from researchers using the database will the OCR reach its potential. Already, numerous research projects using the

OCR have resulted in improvement. The innovative method of cancer registration used in Ontario has resulted in a cancer registry of good quality, particularly in later years where the proportion of cases identified by death certificates is low and the proportion histologically verified is more than 85%. As more computers are introduced into different aspects of the health care system, this unique approach may prove to be the optimal technique for cancer registration in the future.